Thank you Suzi for your introduction. Welcome everyone to the webinar today. My name is Fen Zhao, I'm the Staff Associate, Strategic Innovation at NSF in the Directorate for Computer and Information Sciences and Engineering, in the Office of the Assistant Director. Today I'll be giving you some background on the New BDHubs program.

Before we start, I want to make sure to acknowledge the whole team here at NSF that has been working on the BDHubs program, including the solicitation and charrette series, including the hard work of our intrepid AAAS Fellows Martin Wiener and Al Suarez, program officers and Sylvia Spengler and Chaitan Baru, and the guidance of Suzi Iacono.

In addition I'd want to thank the team of people outside of NSF hard at work hosting the BDHubs charrettes, specifically Casey Henderson, Camille Mulligan, and Natalie DeJarlais of USENIX and Andy Burnett at KnowInnovation, who have all worked real magic with our tight timelines.

Finally I want to thank the steering committee for the Big Data Charrettes for lending their expertise, wisdom, and time to working with NSF to hone the BDHubs charrettes; many thanks to Dan Akins, Fran Berman, Alan Blatecky, David Logsdon, Chris Mentzel, Glenn Ricart, and George Strawn.

The agenda for our webinar today is three fold. Suzi's mentioned a bit about where the broader background of BDHubs came from and why it's a key component to not only NSF, but the broader National Big Data R&D initiative and the federal government. I'll run through some the historical timeline and background leading up to this new program first, setting some context on where this program sits in the broader efforts NSF and other federal agencies have been involved in over the last few years.

Second, I'll try to explain some of the overall strategy and motivation behind this BDHubs program. In short, the multiphase BDHubs program aims to build regionally focused consortia around the country that will ideate, plan, and support Big Data partnerships and collaborative activities. Each BD Hub would have members from academia, industry, non-profits, and government. There will be four Hubs across distinct geographic regions of the United States, including the Northeast, Midwest, South, and West. Each BD Hub should focus on key Big Data challenges and opportunities for its region of service and aim to support the breadth of interested local stakeholders within their respective regions. Ultimately, the BD Hub aims to achieve some common Big Data goals for stakeholders in their region that would not be possible for the independent members to achieve acting alone.

Finally, I'll conclude with some details about the recently released solicitation BDHubs and the series of workshops (called "charrettes") that NSF is sponsoring and hosted by USENIX and KnowInnovation, and how they will support the Big Data Regional Innovation Hubs.

The Big Data Regional Innovation Hubs is just one component of the National Big Data R&D Initiative, first launched in March of 2012 by the White House Office of Science and Technology Policy. To launch the initiative, six Federal departments and agencies today announced more than $200 million in new commitments that, together, promise to greatly improve the tools and techniques needed to access, organize, and glean discoveries from huge

volumes of digital data.

This first phase of the initiative focused primarily on what agencies could do to directly fund more R&D in Big Data. Beginning in the second year of the National Big Data Initiative, the Administration began encouraging multiple stakeholders including federal agencies, private industry, academia, state and local government, non-profits, and foundations, to develop and participate in Big Data innovation projects across the country. In May of 2013, a workshop at the White House brought together such Big Data stakeholders to learn about existing Big Data partnerships, make connections with interested parties, and explore future possibilities. In November of 2014, 29 new partnerships between over 100 different organization were announced and launched at the Data to Knowledge to Action event. Since that event, these partnerships have progressed and many have be able to share their outcomes and results (you can find information on these partnerships and updates on the NITRD Big Data Senior Steering Group webpage).

Today we stand at the threshold of what could be considered the third phase of the National Big Data R&D Initiative, where we take the energy and brilliant ideas shared by Big Data partners during Data2Action, and look to the future in creating a long standing mechanism to drive innovation and partnership in Big Data. The BDHubs programs looks to sustain the level of activity first started with Data2Action, and continue and grow the number and scale of partnership around the nation.

A question many of you in the audience today might have about the BDHubs program is "Why partnerships?" Many of you in the academic community know NSF as a primary funder of research, infrastructure, and education. So where does this motivation to fund more partnerships in Big Data come from? It seems rather novel and different, so I'm going spend a little bit of time explaining why the idea of partnerships is critical to what we see as the health of Big Data innovation in the country.

Dr. John Holdren, the President's Science Advisor and Director of the Office of Science and Technology Policy at the White House said it best back in November at the Data2Action event. He said "America is rich with institutions that are expert at generating data, but as a Nation we have not fulfilled our potential to make the most of these data by merging pre-competitive resources, partnering on analytics, and sharing lessons learned. Today's announcements [at the Data2Action event] show that we are maturing in this respect, finding synergies and collaborative opportunities that will accelerate progress in a wide range of scientific, social, and economic domains."

As we think about what is necessary to bring forth Big Data innovation, to take the capabilities that are available today or on the horizon and turn that into impactful societal change – to enhance national priority areas like economic growth and job creation, education and health, energy and sustainability, public safety and national security, and global development– we begin to realize that it is an impossible task to take on alone. Every organization owns a different piece of the Big Data pie, and have their own purview and expertise and capabilities. By bringing stakeholder from across sectoral divides, we envision that the progress made in research, development, adoption, use, and integration into society could be accelerated to faster pace, be conducted more conscientiously, and ultimately bring deeper permanent impact to society at large.

So before we move forward, I want to make sure we're all on the same page and understand a bit of the terminology we're using today. First I want to explain this term "partnership" that we are using by giving some examples from partnerships that were launched during the Data2Action event. These are just a few highlights from the event that I chose for pedagogical reasons, there were many more excellent partnerships launched at that time.

First on the foundation research, we have Berkeley AMPLab, which is a research lab is funded by government agencies such as NSF, DARPA, DOE. It's also sponsored by a large number of private sector companies including founding partners AWS, Google, and SAP. At the time of Data2Action, the funding between public and private was about 50/50. As an output of their academic foundation research in Big Data, AMPLab creates Apache open source software in their platform, the Berkley Big Data Analytics Stack (BDAS) including Spark/Shark, Mesos, Tachyon. This great software winds up becoming a resource for the entire broader community. In addition, industry sponsors get the added benefit of to interact with researchers and students at various lab meetings, hearing about their progress in cutting edge research and having access to some of the top students in the field that will become the superstars in the Big Data workforce in coming years.

On that note, I want to bring up a second example from the education and workforce component of Big Data, with the Data Science for Social Good summer program hosted by University of Chicago and sponsored by the Schmidt Family Foundation. This program guides fellows (undergraduate and graduate) from multiple academic disciplines through Data Science projects scoped by project partners like City of Chicago, or the Cook Country Sheriff, or the Environment Defense Fund. The fellows produce analysis and applications that will actually be used by project partners, so fellows get the benefit of the experience and partners get an end product that's useful to them.

Finally I wanted to bring up a great example of a partnership in a critical national priority area– healthcare. At Data2Action, Novartis, Eli Lilly, and Pfizer announced their intent to improve access to information about clinical trials by creating a new platform for people running clinical trials and those trying to match clinical trials to patients (so this includes healthcare providers, particularly those developing software for health). This platform builds on a federal open date platform clinicaltrials.gov.

To help regions and stakeholders find each other to form these partnerships that we have mentioned previously, through BDHubs NSF aims to support though developing a coordination network. The top-level of this network are the four geographically based hubs– the basis of what this webinar is about. On can think of each hub having a series of "spoke" which are the local priority areas that the Hub wants to engage in– for example transportation, or manufacturing, or land use–effectively the Big Data application areas. These are the "spokes" of the hub. Each spoke could lead to multiple nodes. These are defined as partnerships between two or more organizations that are geared towards driving some specific end goal in that spoke. We would hope that the hubs during their first award period of 3 years covered by this first solicitation would be able to drive 20-30 of these partnership nodes.

Now, a little more on what the Hubs could do. Each BD Hub should focus on key challenges and opportunities in its region of service. Opportunities could include facilitating partnerships on

overarching themes (e.g., privacy, data sharing, data stewardship, etc.), providing shared resources to the community (e.g., tools, infrastructure, testbeds, etc.) and/or coalescing around key topical themes (e.g., energy, transportation, healthcare). Potential activities for BD Hubs include, but are not limited to:

- Accelerating the ideation and development of Big Data solutions to specific regional and/or societal challenges by convening stakeholders across sectors to partner in results-driven programs and projects;
- Driving successful pilot programs for emerging Big Data technology by acting as a matchmaker between the various academic, industry, and community stakeholders;
- Engaging stakeholders across the region, based on shared interests and industry sector engagement, to enable dialogue and share best practices, and to set standards for data access, data formats, metadata, etc.;
- Increasing the speed and volume of technology transfer between universities, public and private research centers and laboratories, large enterprises, and small and medium-sized businesses;
- Providing data resources of critical importance to the region such as a data steward or public trust service that validates and certifies privacy-sensitive data sets, infrastructure or other testbeds for small scale experimentation, and/or data tools relevant to the analysis needs of stakeholders; and
- Facilitating engagement with opinion and thought leaders on the societal impact of Big Data technologies as to maximize positive outcomes of adoption while reducing unwanted consequences. Topical examples could include privacy or broadening participation.

There are a lot of benefits to Big Data stakeholders to think about partnering with other organizations, and the benefits tend to be different whether the organization is academic or industry or government, and on whether they are data providers, users, big organization and small organizations. I'm going to list a few of the top line broad benefits for why your organization might want to participate in this program.

The first advantage is of course the benefit of initiating new partnership and collective ideation. By interacting with other organizations, your organization will have access to a wider set of ideas and collective action, and getting help for your organization to pursue it's own agenda or mission. One might say that it parallels of the benefits of "Big Data" and all the things that are made possible by merging disparate data sources!

The other is the benefit of working to pool resources and shared infrastructure that individual organizations might not have access to typically or scale to obtain by themselves. User will be able to leverage the resource contributed by other partners, and providers will be able to find users to test out or develop novel applications for that infrastructure.

A third point is connecting organization to facilitate better access to talent through our academic partners. As it is often pointed out, a data scientist is as rare as a unicorn– a mythical being that is often dreamed about but near impossible to find. We would hope that hub partners might be able to not only find some of these unicorns within the schools participating, but perhaps through specific partnership,

give students training and experience while they are still in school working on specific problems that the partner might want to hire for.

A fourth benefit is around sharing best practices and similarly, working on standards. There are many topics where sharing best practices would benefit the community, but one particular one that has strong resonance in recent months is around privacy, discrimination, and the social and ethical implications of Big Data. People involved in data analytics everywhere are concerned and really still trying to sort the basics of what to do in these areas. So providing that healthy discussion and sharing of what we discover works and doesn't work should help ensure adoption of technology while minimizing the unwanted consequences.

Finally the big advantage of NSF's role in the Big Data Hubs program is that we've taken on funding a lot of the coordination costs that always comes with forming partnerships. As most people who have done this before know, there's a very real dollars and cents cost to partnering with another group coming from just the logistics and planning. Through the program, NSF will fund the full time FTEs that will help organize the meeting, and we'll fund the workshops and travel and other costs needed to get the partners together working efficiently.

BDHubs is just one component of a broader Big Data portfolio of programs that NSF funds. Here at NSF and used by the NITRD Big Data Senior Steering Group, we have a four part framework of how we think about Big Data. One is the foundational research that's needed, which at NSF is funded primarily by the Critical Techniques and Technologies for Big Data program (often just called BIGDATA, all caps). This program currently has an open call and I encourage everyone to take a look. We also think about supporting infrastructure design and development, which is primarily supported by the Data Infrastrcutre Building Blocks or DIBBS program. We also fund education in Big Data through multiple programs, but last year and this year, the National Research Traineeship is a very large program with a focus on Data Enabled Science and Engineering. Finally, on the partnerships and engagement quadrant of the portfolio, we have the BDHubs program.

So as you listen through this webinar, you might think "Oh wait this program isn't exactly what I need in terms of funding support…" I encourage you to take a look at the other programs we mention today. Specifically, Chaitan Baru, the program officer for the BIGDATA program will be joining at the end of this webinar and sharing some information about that program.

So as we described the BDHubs program, those experienced with NSF amongst the audience will have probably realized that the BDHubs program is quite different than the majority of NSF programs. Here are some of the big differences that it's important to keep in mind as you think about BDHubs.

First, understand that this solicitation is the first of a multi-phase process meant to develop a whole National Network of BD Hubs. The first phase will set up the governance structure of each BD Hub's consortium of members as well as develop approaches to ensure cross-hub collaboration and sustainability over the long term. The next phase will focus on building out the spokes– the various areas of particular interest to each BD Hub (e.g., transportation, smart cities, health, energy, public safety, and

education). The final phases will focus on connecting the BD Hubs and their regional sectors into a national Big Data innovation ecosystem.

It's important to realize that we want to foster collaboration instead of competition. NSF is hoping that stakeholder in the region can gather and present a cohesive proposal for their region that describes the general consensus. The "lead institution" that submits that proposal is really just a logistical facilitator for the Hub.

Also, we expect that while stakeholders will have a general sense of some of the initial areas of interest and participating organizations for the Hub when they develop their proposal to NSF, the Hubs will be dynamic and grow over time to include more participants. The solicitation and charrette series we describe later will only be the kickoff to the Hubs program; in the future there will be many more opportunities for your organization to engage in BDHubs activities.

Finally, I just want to reiterate that the funding from the BDHubs program will be for staff and networking activities (like workshops), not research. Now, one could foresee that while partners work together they might develop some ideas for research that would be appropriate for NSF to fund– but those partners would have to go to other programs at NSF like BIGDATA or DIBBS to get support for those activities, and not through this set of calls for proposals.

And on that note, I'd like to move onto some specifics of the solicitation that was posted as NSF document 15-562. This solicitation is just the first phase of the BDHubs program, with a primary focus on setting up the organizational infrastructure and governance of the regional hubs.

NSF expect to fund 4 awards (one in each region). Each award will be a maximum of $1.25M over 3 years, subject to availability of funds. Therefore the anticipated total funding amount will be $4M to $5M. The due date for the proposals is June 24th 2015.

Proposals can be submitted by a wide number of organizations such as universities, colleges, non-profits, state and local governments, as well as national labs and FFRDCs. NSF welcomes collaborative proposals from for-profit organizations as well, but those organization can only be subawardees and not the lead awardee. An organization can submit only one proposal to this solicitation (i.e. across all the regions), and an individual can only serve as PI or Co-PI in at most one submission.

NSF is not prescriptively defining how a regional hub should be structured; we're letting the regional stakeholders decide on what kind of specific structure would work best. We're only giving a guideline that the Hub should have a broad 3 part structure to the hub.

There will be a steering council tasked with making key decisions and setting the agenda for the regional consortium as a whole. The council will be made up of unpaid representatives from a subset of participating organization, and should be representative of the Hub's membership (e.g. across sectors, states, areas of interest), while also considering participation from traditionally underrepresented groups.

There will also be a full time paid executive director and associated staff that the proposing institution will pay with NSF funds. The staff will implement the decisions of the steering council and oversee the day to day operations of the Hub.

Finally there will the partner organizations that make up the "nodes" of the hub. An organization doesn't need to have a representative in the steering council to be a partner organization– in fact they should only send a representative to the steering council if they have an interest in the broader governance of the hub rather than any specific activity or priority area. Note that if an organization does not need to be located within a region to be a partner in a node since many organizations have a national scope.

A little more on the regional definitions for this program. For the purposes of this program, we're relying on the US Census definition of the Northeast, Midwest, West, and South. Alaska and Hawaii are part of the West Region. Organizations based in US territories can choose the region they would like to participate in. There will be some larger organizations, particularly in the industry sector, that have national scope and a substantive presence in multiple regions. Those organization should choose to join the regional Hub most appropriate for their participation, and will be limited to being represented on the steering council of at most one regional hub. However they can be partners on specific nodes of potentially multiple regions.

In the proposal to NSF, we've asked for details about how the hubs will be structured, what needs to be funded, what preliminary set of areas and partnerships the Hub would be interested, as well a general idea of how the Hub might be made sustainable in the future. Rather than repeated the details from the solicitation, I encourage everyone to take a close look at the proposal preparation section of the solicitation text.

The review process for the solicitation will be by the standard NSF merit review process with a panel and additional ad-hoc reviews. As this is a unique solicitation, we've added a few additional review criteria to the standard intellectual merit and broader impacts criteria.

Generally proposals submitted in response to this BD Hubs solicitation should be from organizations that are currently engaged in Big Data innovation activities. These organizations should have a history of leading or fostering collaborations among multiple Big Data stakeholders and must be ready to build further partnerships. Also, BD Hub proposals should describe explicit mechanisms for creating and fostering collaborations in their regions. Vehicles for such collaboration may include creation of seminar series (with ability for remote presentations and participation); regional working groups; and "visitor" and other exchange programs to cross-pollinate collaborations and activities across organizations.

Therefore the two affiliated review criteria are around how well the proposal addresses the challenges of opportunities of the region, specifically around those that are not possible for single organizations to tackle. Also we ask to judge how well the Hub will work to include the breadth of stakeholders in the region.

Additionally, each proposal must also foster efforts to educate and train the Big Data workforce in its region, such as data scientists, business managers, students, and end users. Example efforts could include, but are not limited to, new data science curricula, educational workshops open to the region, and hackathons or other competitions focusing on one or more Big Data challenge(s).

Finally, fostering of a sustainable innovation ecosystem is critical to a successful BD Hub. Each proposal must therefore address the sustainability of the BD Hub in the long term.

To close, I would like to have a discussion around the upcoming charrettes for the BDHubs program. A charrette is an intensive design session. These 1 day events will help kick off the process of stakeholders from within regions coming together and discussing how a Hub might be designed to best address their interests and benefit them.

The four charrettes are scheduled to start next week:

- April 8th, 2015: Midwest, in Ann Arbor, MI (Sheraton Ann Arbor Hotel);
- April 10th, 2015: West, in Salt Lake City, UT (Sheraton Salt Lake City Hotel);
- April 13th, 2015: South, in Durham, NC (Marriott at Research Triangle Park);
- April 17th, 2015: Northeast, in Boston, MA (Four Points by Sheraton Wakefield Boston Hotel & Conference Center).

These charrettes are being sponsored by NSF, hosted by the USENIX Association, and facilitated by KnowInnovation.

As I mentioned, each *charrette* is meant to be an intensive, one-day design and planning workshop with the objective of convening stakeholders in that region around a common set of Big Data challenges — particularly those that may be especially relevant to that region. Each charrette aims to help establish a regional consortium that builds upon existing efforts within the region. These meetings bring together academic, non-profit, governmental, and business communities throughout the country to form grassroots regional partnerships to foster and propel Big Data approaches across all sectors. These communities represent stakeholders in the Big Data ecosystem, including corporations, universities, philanthropies, non-profits, and state and local governments.

In each charrette, attendees will engage in a discussion on how to form and sustain a partnership within their region. Each attendee is expected to represent an organization engaged in Big Data activities in that region, and should be able to think strategically about his/her organization's decisions regarding partnerships and regional collaborations. The charrette is meant to convene stakeholders to collaboratively form a single consortium for their region. Attendees will discuss which organization could take on the coordinating role of establishing the regional consortium. Participants should be willing to engage in frank discussion and assessment of ideas in a collegial and professional fashion. USENIX, in partnership with Knowinnovation, will assemble a team of mentors and provocateurs, selected for their relevant expertise, as well as professional facilitators to aid the participants in the discussion of topics and development of ideas.

The charrettes are really meant to be a first step in the process of gathering participants in designing a regional hub by offering valuable face to face time for interested partners to meet each other and discuss potential avenues of future work together. If your organization can't sent someone to attend the charrette, NSF has set up a virtual forum at bdhub.info using the HUBzero platform. At bdhub.info, you'll be able to join the "Group" for your region. After the charrettes, we'll ask the attendees of the charrettes to use the platform to continue lines of communication and set up future meetings within the region. Let me re-emphasize here that there will be multiple future opportunities to engage in BDHubs activities.

So I think we're going to have very exciting time at the charrettes coming up and I hope to see many of you there. Here is my contact information, and feel free to reach out to me if you have any questions. We'll open up the line to take some questions. I want to add that joining us to answer some questions are Chaitan Baru, the lead of the NSF BIGDATA program. Additionally we have Andy Burnett, the lead at KnowInnovation who can answer any questions about the activities to happen at the upcoming charrettes.

As we wait for questions to come in, here's a first one that I know is on many of your minds. This is one Chaitan Baru can best answer– "I've got a great collaborative research project was wondering if the BDHubs program could fund this?"